

Strong noise estimation in cubic splines

A. Dermoune^a, A. El Kaabouchi^b

^a*Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524. UFR de Mathématiques,
Bât. M2, 59655 Villeneuve d'Ascq Cedex, France*

^b*Institut Supérieur des Matériaux et Mécaniques Avancés, 44, Avenue Bartholdi,
72000, Le Mans, France*

Abstract

The data $(y_i, x_i) \in \mathbb{R} \times [a, b], i = 1, \dots, n$ satisfy $y_i = s(x_i) + e_i$ where s belongs to the set of cubic splines. The unknown noises (e_i) are such that $\text{var}(e_I) = 1$ for some $I \in \{1, \dots, n\}$ and $\text{var}(e_i) = \sigma^2$ for $i \neq I$. We suppose that the most important noise is e_I , i.e. the ratio $r_I = \frac{1}{\sigma^2}$ is larger than one. If the ratio r_I is large, then we show, for all smoothing parameter, that the penalized least squares estimator of the B -spline basis recovers exactly the position I and the sign of the most important noise e_I .

Key words: B-spline functions, Cubic splines, hat matrix, Moore-Penrose pseudoinverse.

1991 MSC: 2000, 15A09, 41A15, 62J05.

1 Linear inverse problem: General setting

The data $(y_i, x_i) \in \mathbb{R} \times [a, b], i = 1, \dots, n$ satisfy

$$y_i = s(x_i) + e_i. \quad (1.1)$$

The map $s : [a, b] \rightarrow \mathbb{R}$ is unknown, and (e_i) are the error of measurements, also called the noise and is unknown. We suppose that s belongs to a set \mathcal{C} of functions, and we are interested in the estimation of $s \in \mathcal{C}$ using the data $(y_i, x_i), i = 1, \dots, n$. Suppose that \mathcal{C} has a basis $(b_j)_{j=1, \dots, d}$. In this case, each map $s \in \mathcal{C}$ is determined by its coordinates $\beta = (\beta_1 \dots \beta_d)^T$ in the latter basis,

Email addresses: Azzouz.Dermoune@math.univ-lille1.fr (A. Dermoune), aek@ismans.fr (A. El Kaabouchi).

i.e., $\forall x \in [a, b]$, $s(x) = \sum_{j=1}^d \beta_j b_j(x)$. Hence, for each i , $s(x_i) = \mathbf{b}(x_i)\boldsymbol{\beta}$ with $\mathbf{b}(x_i) = (b_1(x_i) \dots b_d(x_i))$ is the $(1, d)$ matrix. If we introduce the (n, d) matrix

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}(x_1) \\ \vdots \\ \mathbf{b}(x_n) \end{pmatrix}, \quad (1.2)$$

then, the data $\mathbf{y} = (y_1 \dots y_n)^T$ and the noise $\mathbf{e} = (e_1 \dots e_n)^T$ satisfy the linear system

$$\mathbf{y} = \mathbf{B}\boldsymbol{\beta} + \mathbf{e}. \quad (1.3)$$

The latter is known as the linear regression in Statistic community and the linear inverse problem in the Inverse problem community. This problem is ill-posed, because the transformation $\boldsymbol{\beta} \mapsto \mathbf{B}\boldsymbol{\beta}$ is not invertible. Moreover, the noise \mathbf{e} is not known.

One way to estimate the parameter $\boldsymbol{\beta}$ and the noise \mathbf{e} is to use the generalized penalized least square estimators. It works as following. We fix a matrix \mathbf{M} having n columns, and we consider the ellipsoide quasi-norm $\|\cdot\|_{\mathbf{M}}$ defined by $\|x\|_{\mathbf{M}}^2 = x^T \mathbf{M}^T \mathbf{M} x$. We propose, for each $\lambda > 0$ and for each matrix \mathbf{L} having d columns, the minimizers

$$\hat{\boldsymbol{\beta}}(\lambda, \mathbf{M}, \mathbf{L}) \in \arg \min_{\boldsymbol{\beta}} \{\|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|_{\mathbf{M}}^2 + \lambda \|\mathbf{L}\boldsymbol{\beta}\|^2\} \quad (1.4)$$

as an estimator of the vector $\boldsymbol{\beta}$. The quantity $\|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|_{\mathbf{M}}^2$ is the square of the residual error with respect to the metric defined by the quasi-norm $\|\cdot\|_{\mathbf{M}}$, and $\|\mathbf{L}\boldsymbol{\beta}\|^2$ is called the penalty. The parameter λ is called the smoothing parameter. We have easily the following results.

Proposition 1 *The set of the minimizers of the latter optimization is given by the following normal equation*

$$(\mathbf{B}^T \mathbf{M}^T \mathbf{M} \mathbf{B} + \lambda \mathbf{L}^T \mathbf{L})\boldsymbol{\beta} = \mathbf{B}^T \mathbf{M}^T \mathbf{M} \mathbf{y}. \quad (1.5)$$

If $N(\mathbf{M}\mathbf{B}) \cap N(\mathbf{L}) = \{0\}$, then $\hat{\boldsymbol{\beta}}(\lambda, \mathbf{M}, \mathbf{L})$ is unique and is given by

$$\hat{\boldsymbol{\beta}}(\lambda, \mathbf{M}, \mathbf{L}) = (\mathbf{B}^T \mathbf{M}^T \mathbf{M} \mathbf{B} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{B}^T \mathbf{M}^T \mathbf{M} \mathbf{y} := \mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}. \quad (1.6)$$

Here, $N(\mathbf{A})$ denotes the null-space of the matrix \mathbf{A} .

The minimizer $\mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}$ is proposed as an estimator of the parameter $\boldsymbol{\beta}$. Hence, $\mathbf{B}\mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}$ is an estimator of $\mathbf{B}\boldsymbol{\beta}$ and $\mathbf{y} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}$ is an estimator of the noise \mathbf{e} . The map $x \in [a, b] \mapsto \sum_{j=1}^d \mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}(j)b_j(x)$ is an

estimator of the map s . The performance of these estimators depends clearly on the matrix $\mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})$ known as the hat matrix in Statistics community.

Proposition 2 *The limit $\lim_{\lambda \rightarrow 0^+} \mathbf{H}(\lambda, \mathbf{M}, \mathbf{L}) := H(0, \mathbf{M}, \mathbf{L})$ exists and is equal to the \mathbf{ML} -weighted pseudoinverse of \mathbf{B} defined by*

$$\mathbf{B}_{\mathbf{ML}}^+ = (\mathbf{I} - (\mathbf{LP}_{\mathbf{MB}})^+ \mathbf{L})(\mathbf{MB})^+ \mathbf{M}. \quad (1.7)$$

Here \mathbf{A}^+ denotes the Moore-Penrose inverse of \mathbf{A} and

$$\mathbf{P}_{\mathbf{A}} = \mathbf{I} - \mathbf{A}^+ \mathbf{A} \quad (1.8)$$

denotes the orthogonal projection on $N(\mathbf{A})$.

It follows that the estimator $\hat{\beta}(\lambda, \mathbf{M}, \mathbf{L})$ converges to $\mathbf{B}_{\mathbf{ML}}^+ \mathbf{y} := \hat{\beta}(0, \mathbf{M}, \mathbf{L})$ as $\lambda \mapsto 0$. Moreover, we can show that

$$\hat{\beta}(0, \mathbf{M}, \mathbf{L}) = \arg \min_{\beta} \{ \|\mathbf{L}\beta\|^2 : \beta \in \arg \min \|\mathbf{y} - \mathbf{B}\beta\|_{\mathbf{M}}^2 \}. \quad (1.9)$$

In particular, if \mathbf{B} has maximal rank, i.e., $R(\mathbf{B}) = \mathbf{R}^n$, then

$$\hat{\beta}(0, \mathbf{M}, \mathbf{L}) = \arg \min_{\beta} \{ \|\mathbf{L}\beta\|^2 : \mathbf{y} = \mathbf{B}\beta \}, \quad (1.10)$$

or equivalently

$$\mathbf{B}\mathbf{B}_{\mathbf{ML}}^+ \mathbf{B} = \mathbf{B}. \quad (1.11)$$

Observe that, if \mathbf{M} is invertible, then

$$\mathbf{B}_{\mathbf{ML}}^+ \mathbf{B} = \mathbf{B}_{\mathbf{L}}^+ \mathbf{B}. \quad (1.12)$$

On the other hand, the limit

$$\lim_{\lambda \mapsto +\infty} (\mathbf{B}^T \mathbf{M}^T \mathbf{M} \mathbf{B} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{B}^T \mathbf{M}^T \mathbf{M} \quad (1.13)$$

exists and is equal to

$$\mathbf{C}_{\mathbf{MBL}} := (\mathbf{MBP}_{\mathbf{L}})^+ \mathbf{M}. \quad (1.14)$$

In the classical case $\mathbf{M} = \mathbf{I}$ and $\mathbf{L} = \mathbf{I}$, the matrix

$$\mathbf{H}(\lambda, \mathbf{I}, \mathbf{I}) = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{I})^{-1} \mathbf{B}^T \quad (1.15)$$

is known as the Tikhonov regularization of the Moore-Penrose inverse of the matrix \mathbf{B} .

PROOF. It is a consequence of known results[2].

In the sequel, we suppose that the noise \mathbf{e} is Gaussian with the covariance matrix $\mathbf{C} = \text{diag}(\sigma_i^2)$. In this case, the natural choice of \mathbf{M} is the weight

matrix $\mathbf{C}^{-1/2}$. We suppose that the variance $\sigma_I^2 = 1$ for some I and $\sigma_i^2 = \sigma^2$ for all $i \neq I$. The set of functions \mathcal{C} is the set of cubic splines. The true signal is an element of the B -spline basis. We consider, for each $\lambda > 0$, the noise estimator $\hat{\mathbf{e}}(\lambda) = \mathbf{y} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}$. The aim of our work is to show that if σ^2 is small, then $\hat{\mathbf{e}}(\lambda)$ recovers exactly the position and the sign of the most important noise. Section 2 recalls some cubic splines results. In Section 3 we present our numerical results.

2 Cubic splines

Schoenberg introduced in [6] the terminology spline for a certain type of piecewise polynomial interpolant. The ideas have their roots in the aircraft and shipbuilding industries. Since that time, splines have been shown to be applicable and effective for a large number tasks in interpolation and approximation. Various aspect of splines and their applications can be found in [3]. Let $a = \kappa_0 < \kappa_1 < \dots < \kappa_{K+1} = b$ be a sequence of increasing real numbers. Spline interpolation can be described as following. A map s belongs to the set $S_3(\kappa_0, \dots, \kappa_{K+1})$ of cubic splines with knots $(\kappa_0 < \kappa_1, \dots, \kappa_{K+1})$ if

$$s(x) = p_i + q_i(x - \kappa_i) + \frac{u_i}{2}(x - \kappa_i)^2 + \frac{v_i}{6}(x - \kappa_i)^3 \quad (2.1)$$

for every $x \in [\kappa_i, \kappa_{i+1})$. Let $(b_j := S_{j,4} : j = -3, \dots, K)$ be the B -spline basis functions of the set $S_3(\kappa_0, \dots, \kappa_{K+1})$.

Before going further, let us recall the famous result of [6] and [5]. If the data $(y_i, \kappa_i) : i = 0, \dots, K+1$, then the minimizer of

$$\min_{s \in S_3(\kappa_0, \dots, \kappa_{K+1})} \left\{ \lambda \int_a^b |s^{(2)}(x)|^2 dx + \sum_{i=0}^{K+1} |s(\kappa_i) - y_i|^2 \right\} \quad (2.2)$$

is the natural cubic spline, i.e. such that $s^{(2)}(\kappa_0) = s^{(2)}(\kappa_{K+1}) = 0$, where $s^{(2)}$ is the second derivative of s . Observe that the penalized matrix \mathbf{L} is defined by

$$\|\mathbf{L}\boldsymbol{\beta}\|^2 = \int_a^b |s^{(2)}(x)|^2 dx. \quad (2.3)$$

Let us calculate the matrix \mathbf{L} . The unknown vector $\boldsymbol{\beta}$ belongs to \mathbb{R}^{K+4} . From the derivative formula for B -spline functions [1], ch. X. we have

$$\sum_{j=-3}^K \beta_j S_{j,4}^{(2)}(x) = \sum_{j=-1}^K \beta_j^{(2)} S_{j,2}(x), \quad (2.4)$$

where the vector $\boldsymbol{\beta}^{(2)} = \Delta_2 \boldsymbol{\beta}$ and Δ_2 denotes the matrix corresponding to the weighted difference operator. If we denote by \mathbf{R} the matrix with entries

$$R_{ij} = \int_a^b S_{j,2}(x) S_{i,2}(x) dx, \quad i, j = -1, \dots, K, \quad (2.5)$$

then

$$\int_a^b |S^{(2)}(x) \boldsymbol{\beta}|^2 dx = \boldsymbol{\beta}^t \Delta_2^t \mathbf{R} \Delta_2 \boldsymbol{\beta} = \|\mathbf{L} \boldsymbol{\beta}\|^2, \quad (2.6)$$

with $\mathbf{L} = \mathbf{R}^{1/2} \Delta_2$.

In the sequel we suppose that the data (y_i, x_i) , $i = 1, \dots, n$ with $(x_i : i = 1, \dots, n)$ do not necessarily coincide with the knots $(\kappa_i : i = 0, \dots, K+1)$.

We want to study the estimator $\mathbf{y} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M}, \mathbf{L})\mathbf{y}$ with respect to the smoothing parameter $\lambda > 0$. More precisely we want to recover the position and the sign of the most important noise.

3 Numerical computation

We consider $a = 0$, $b = 1$, $K, n \in \mathbb{N}^*$ and $(\kappa_i)_{i=0, \dots, K+1}$ with $\kappa_{i+1} - \kappa_i = \frac{b-a}{K+1}$ for all $i \in \{0, \dots, K\}$. The data $(y_i, x_i)_{i=1, \dots, n}$ are such that $x_{i+1} - x_i = \frac{b-a}{n-1}$ for all $i \in \{1, \dots, n-1\}$. The model is $\mathbf{y} = \mathbf{B}\boldsymbol{\delta}_j + \mathbf{e}$.

The following show that for all j and for each smoothing parameter λ the noise estimator $[\mathbf{I} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M})]\mathbf{y}$ recovers exactly the position I and the sign of the most important noise. We fix the variance $\sigma^2 \in (0, 1)$, and we consider, for each realization of the noise \mathbf{e} , the maps $I(\mathbf{e})$ and $\text{sgn}(\mathbf{e})$ defined respectively by:

$$\lambda \in (0, 10) \rightarrow \arg \max_i |[\mathbf{I} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M})]\mathbf{y}(i)| = I(\mathbf{e}, \lambda) \in \{1, \dots, n\}, \quad (3.1)$$

$$\lambda \in (0, 10) \rightarrow \text{sgn}(\mathbf{e}, \lambda) = \text{sign}([\mathbf{I} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M})]\mathbf{y}(I)). \quad (3.2)$$

where $\text{sign}(x) = -1$ if $x < 0$, $\text{sign}(x) = 1$ if $x > 0$.

We repeat 100 realizations $(\mathbf{e}^{(k)} : k = 1, \dots, 100)$. We calculate the proportion

$$p_1(\sigma, \lambda, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[I(\mathbf{e}^{(k)}, \lambda) \neq I]}, \quad (3.3)$$

i.e. the probability that the estimator $[\mathbf{I} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M})]\mathbf{y}$ does not recover the position I of the strong noise \mathbf{e}_I . We also calculate the probability that the estimator $[\mathbf{I} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M})]\mathbf{y}$ does not recover the $\text{sign}(\mathbf{e}_I)$ of the strong noise

e_I , i.e.

$$p_2(\sigma, \lambda, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[\text{sgn}(\mathbf{e}^{(k)}, \lambda) \neq \text{sign}(e_I)]}. \quad (3.4)$$

The probability that the path $\lambda \in (0, 10) \rightarrow I(\mathbf{e}^{(k)}, \lambda)$ does not coincide with the position I of the most important noise e_I is equal to

$$p_3(\sigma, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[I(\mathbf{e}^{(k)}) \neq I]}. \quad (3.5)$$

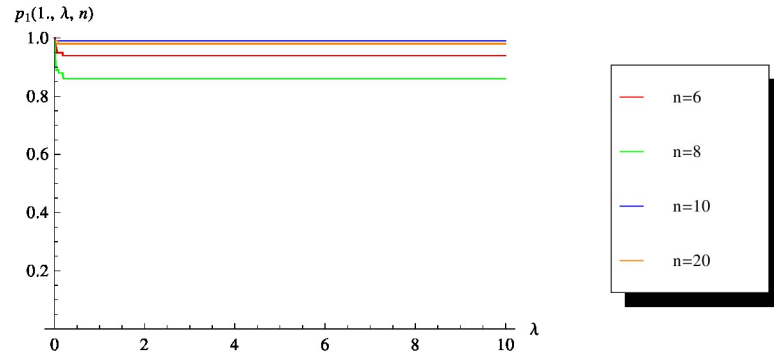
The probability that the path $\lambda \in (0, 10) \rightarrow \text{sign}(\mathbf{e}^{(k)}, \lambda)$ does not coincide with the sign of the most important noise e_I is equal to

$$p_4(\sigma, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[\text{sgn}(\mathbf{e}^{(k)}) \neq \text{sign}(e_I)]}. \quad (3.6)$$

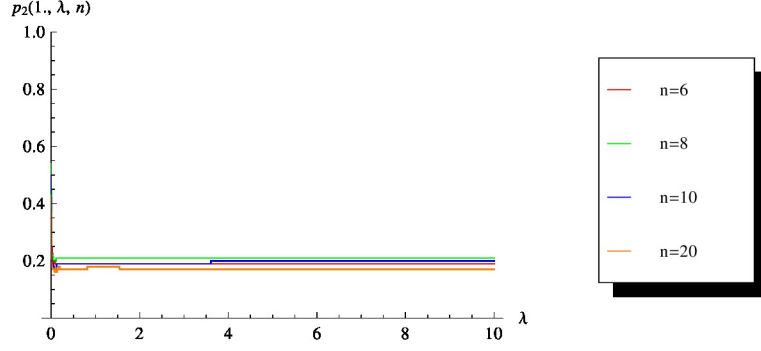
Below we plot $\lambda \in (0, 10) \rightarrow p_l(\sigma, \lambda, n)$ for $l = 1, 2$ and for fixed σ , $\sigma \in (0, 1.5) \rightarrow p_l(\sigma, \lambda, n)$ for $l = 1, 2$ and for fixed λ , and $\sigma \rightarrow p_l(\sigma, n)$ for $l = 3, 4$.

The following example illustrates our results when $K = 4$, $j = 3$, $I = 1$, $n = 6$, $n = 8$, $n = 10$ and $n = 20$.

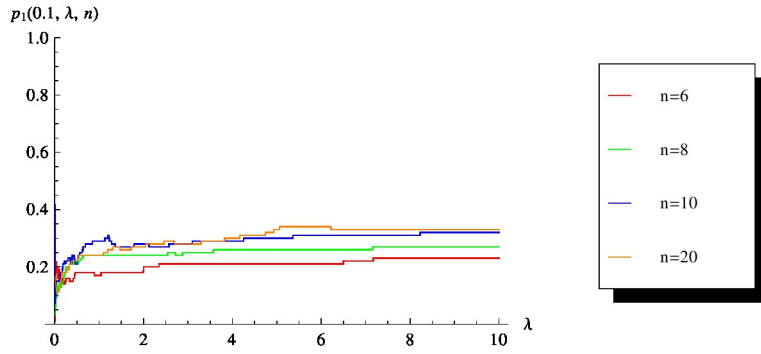
- Plots of $p_1(\sigma, \lambda, n)$ and $p_2(\sigma, \lambda, n)$ for fixed σ .



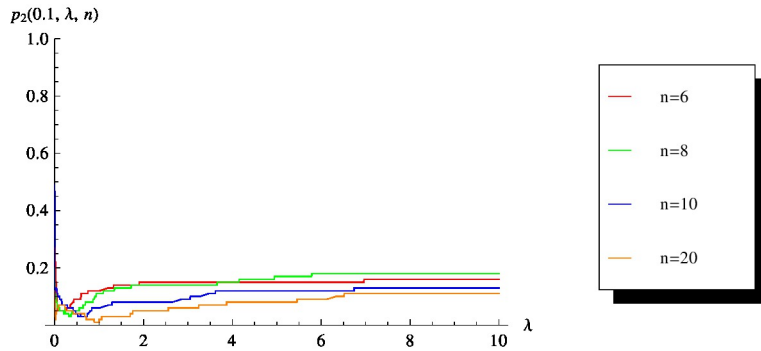
If the noise is white, then there is no dominating noise among e_1, \dots, e_n . Numerical result are coherent. It tells us that the probability that the estimator $[\mathbf{I} - \mathbf{B}\mathbf{H}(\lambda, \mathbf{M})]\mathbf{y}$ recovers the position of the most important noise is very small.



Numerical result shows that even the noise is white the probability that the estimator $[\mathbf{I} - \mathbf{BH}(\lambda, \mathbf{M})]\mathbf{y}$ recovers the sign of the noise is nearly equal to 0.8.



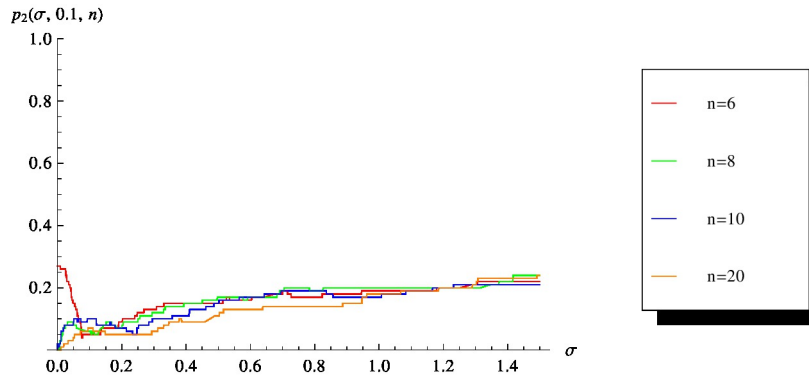
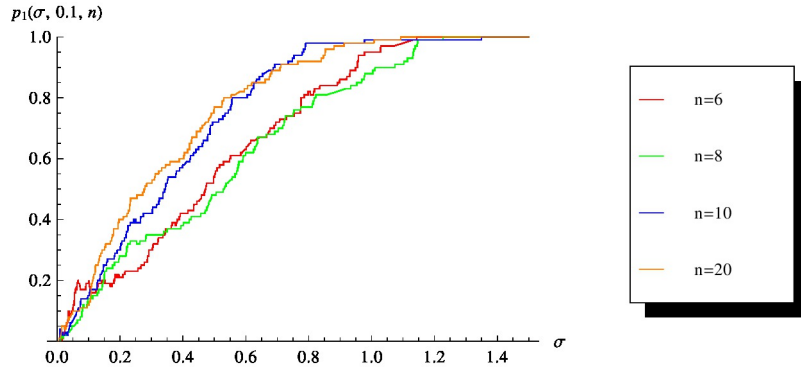
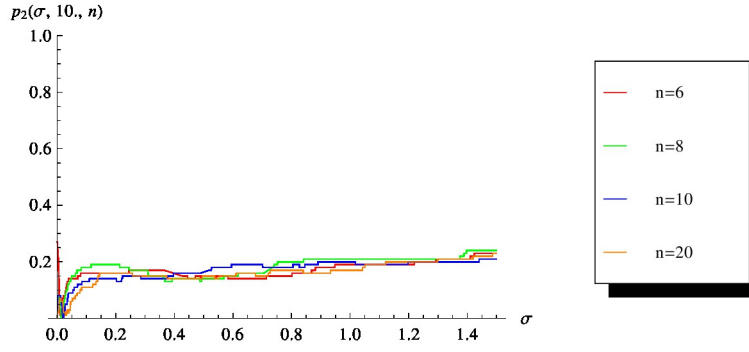
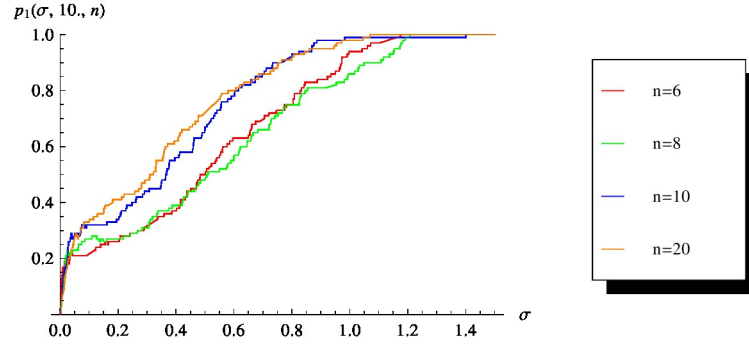
If the noise has a dominating component, then the probability that the estimator $[\mathbf{I} - \mathbf{BH}(\lambda, \mathbf{M})]\mathbf{y}$ recovers the position of the most important noise belongs to $(0.7, 0.9)$ for all $\lambda \in (0, 10)$.



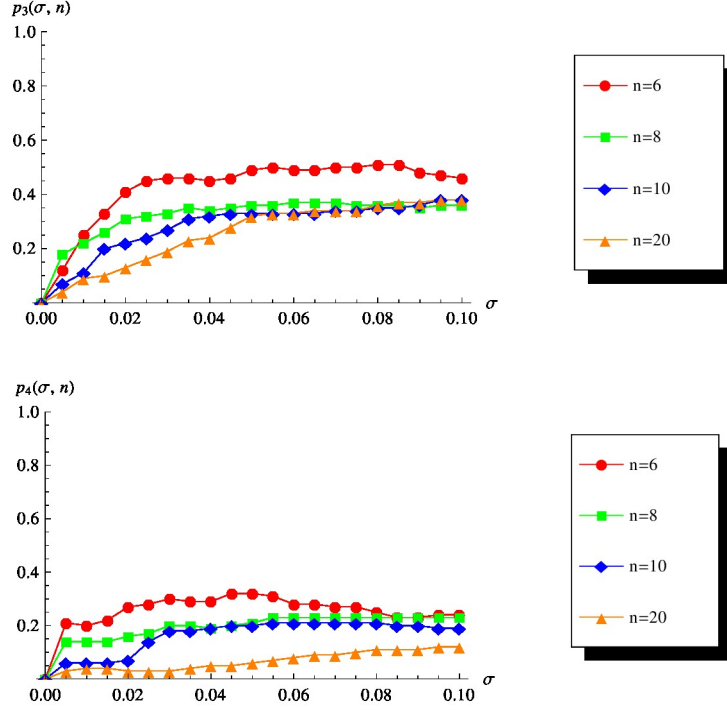
If the noise has a dominating component, then the probability that the estimator $[\mathbf{I} - \mathbf{BH}(\lambda, \mathbf{M})]\mathbf{y}$ recovers the sign of the most important noise belongs to $(0.8, 1)$ for all $\lambda \in (0, 10)$.

- Plots of $\sigma \rightarrow p_1(\sigma, \lambda, n)$ and $\sigma \rightarrow p_2(\sigma, \lambda, n)$ for fixed λ . Numerical results show that both are increasing, but $\sigma \rightarrow p_1(\sigma, \lambda, n)$ increases quickly than

$$\sigma \rightarrow p_2(\sigma, \lambda, n).$$



- Plot of $p_3(\sigma, n)$ and $p_4(\sigma, n)$.



References

- [1] C. de Boor. A Pratical Guide to Splines, New-York, 1978.
- [2] L. Elden. A weighted pseudoinverse, generalized singular values, and constrained least squares problems, BIT, 22 (1982), pp. 487-501.
- [3] Micula G, Micula S (1999): Handbook of splines, Kluwer Academic Publishers, Dordrecht, the Netherlands.
- [4] Hochstenbach, Ninch, Reichel. Discrete ill-posed least-squares problems with a solution norm constraint. Linear Algebra and its Applications (2011). (2007) 935-958.
- [5] C. H. Reinsch. Smoothing by Spline Functions, Numerische Mathematik 10, (1967) 177-183.
- [6] I. J. Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions, Quart. Appl. Math., vol. 4, (1946) 4599 and 112141.